

## Final Technical Report

**Grant #:** N00014-99-1-1032

**Principal Investigator:** Robert K. Atkinson (Email: atkinson@ra.msstate.edu)

**Institution:** Mississippi State University

**Grant Title:** Studies in Computer-Based Learning of Math: Using Self-Explanations and Animated Pedagogical Agents

**Reporting Period:** 26 August 1999 – 31 December 1999 (4 months)

**Award Period:** 26 August 1999 – 31 December 1999

**Objectives:** This project was designed to examine ways of structuring a computer-based learning environment to enhance that learning from examples. In particular, two types of structural manipulations were considered: (1) incorporating an instructional format consisting of instructional explanations and/or self-explanations prompts, a structural manipulation that may have implications for the way in which students study examples; and (2) the use of an animated pedagogical agent, a manipulation that appears to have the potential of optimizing computer-based learning.

**Approach:** In order to address these objectives, a series of lab-based experiments were conducted. The first and third experiments were run at Mississippi State with undergraduates serving as subjects. The second experiment was conducted at State University of New York-College at Oneonta with undergraduates serving as subjects.

**Accomplishments:** Although there is no precise definition of a worked example, it may be described as an instructional device that provides a model for solving a particular type of problem by presenting the solution in a step-by-step fashion. It is intended to provide the learner with an expert's solution, which the learner can use as a model for his or her own problem solving. Moreover, there are a number of practical reasons for using worked examples in instruction: First, students seem to prefer examples over text-based explanations (LeFevre & Dixon, 1986). Second, examples are ubiquitous, as they are found in many textbooks in problem-solving domains (Mayer, Sims, & Tajika, 1995). Third, Silver and Marshall (1990) contend that worked examples "provide an effective mechanism through which learners can become acquainted with a large number of problem solutions much more rapidly than if all of the solutions were generated by the learner" (p. 283). Finally, and perhaps most importantly, worked examples can be effectively employed during the acquisition of the domain-specific knowledge necessary for problem solving (Cooper & Sweller, 1987; Paas & Van Merriënboer, 1994). Despite the fact that worked examples are often recognized as significant sources of learning in knowledge-rich domains such as physics and mathematics, a number of factors seem to account for their effectiveness, including individual differences in example processing and structural features of examples themselves (for review, see Atkinson, Derry, Renkl, & Wortham, in press).

**REPORT DOCUMENTATION PAGE**Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE (DD-MM-YYYY)</b>		<b>2. REPORT DATE</b> Final Technical Report		<b>3. DATES COVERED (From - To)</b> Aug 99 - Aug 00	
<b>4. TITLE AND SUBTITLE</b> Studies in Computer-Based Learning of Math: Using Self-Explanations and Animated Agents				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b> N000149911032	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Atkinson, Robert, K.				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSORING/MONITORING AGENCY REPORT NUMBER</b>	
<b>12. DISTRIBUTION AVAILABILITY STATEMENT</b> Approved for Public Release; distribution is Unlimited.					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b> This study was designed to examine ways of structuring a computer-based learning environment to enhance learning from worked-out examples. In particular, two types of structural manipulations were considered: (1) incorporating an instructional format consisting of instructional explanations and/or self-explanations prompts, a structural manipulation tha may have implications for the way in which students study examples; and (2) the use of an animated pedagogical agent, a manipulation that appears to have the same potential of optimizing computer-based learning.					
<b>15. SUBJECT TERMS</b> Animated Pedagogical Agents, Self-Explanations, Computer-Based Learning Environments					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b>
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			<b>19b. TELEPHONE NUMBER (Include area code)</b>

### Individual Differences in Example Processing

Indeed, research has documented that learners do not employ unidimensional study or example processing strategies when they interact with worked-out examples in physics (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Renkl, 1997). For example, Chi and her colleagues (Chi et al., 1989) noted that physics examples drawn from college textbooks often fail to contain all of the information necessary to solve problems. In particular, examples typically do not include all of the reasons why a certain step in the solution was performed. As a result, the burden of explaining the solution steps rests on the learner. Chi et al. discovered that students in their study attempted to circumvent these sorts of poorly elaborated examples by employing qualitatively different example-studying strategies. The authors observed that some students attempted to establish a rationale for the solution steps by pausing to explain the examples to themselves, and that these students appeared to learn more than those who did not. Chi et al. termed this phenomenon the self-explanation effect.

Chi et al. (1989) also observed rather pronounced differences between more-successful (i.e., better problem solvers) and less-successful students in terms of this self-explanation effect. In particular, they found that the more-successful students differed from the less-successful students in four distinct ways. First, more-successful students uttered a greater number of self-explanations during example studying. Second, the more-successful students voiced a greater number of accurate self-monitoring statements during example studying. Third, while problem solving, the more-successful students were less likely to refer back to the examples. Finally, when more-successful students did take time to look back at the examples during problem solving, their references were more focused.

A recent study by Renkl (1997) examined the issue of whether the quality of self-explanations is a unidimensional construct. Based on his research with college students learning probability calculations from worked examples, Renkl generated a set of profiles of ineffective and effective learners. Consistent with Chi et al.'s (1989) research, he found that performance for both groups was directly related to qualitative differences in self-explanation characteristics. More importantly, Renkl (1997) discovered that the quality of self-explanations among individuals was not unidimensional in nature, although that quality appeared to be a stable trait within individuals. In fact, he discovered four relatively distinct self-explanation styles, two associated with successful problem-solving strategies and two associated with inadequate strategies. Successful students could be classified as either anticipative reasoners or principle-based explainers. Anticipative reasoners tended to self-explain by anticipating the next step in an example solution, then checking to determine whether the predicted step corresponded to the actual step. Principle-based explainers, on the other hand, tended to identify the essential meaning of a problem by attempting to articulate its goal structure, including the application of operators, while also elaborating upon the principle that the example was intended to convey. Although a learner characterized by either style could be described as a "good self-explainer," as Renkl points out, this does not necessarily mean that the individual was good at all facets of self-explanation.

### Structural Features of Examples that Influence Example Processing

Although researchers (e.g., Catrambone, 1994; Sweller, 1993; Ward & Sweller, 1990) acknowledge that the way in which worked examples are structured by instructional designers

plays a critical role in their effectiveness, little research has been dedicated to identifying ways of improving example structure, or to exploring design techniques that enhance example effectiveness, especially within computer-based learning environments. To date, most experiments have utilized worked examples that are visually fixed in nature; that is, the examples simultaneously present a problem and an expert's solution steps. As such, these worked examples are similar to those found in traditional mathematics and science texts; however, instructional materials delivered on multimedia computer systems need not be limited in this way. In fact, when examined together, a number of recent articles hint at ways in which instructional examples could be structured within a computer-based multimedia environment so that learning is maximized. These include sequentially presenting problem states, emphasizing conceptually-related problem steps by visually isolating and labeling solution steps, as well as incorporating a second modality.

In a recent study, Stark (1999) examined whether computer-based examples could be structured so that they encourage learners to engage in anticipative reasoning, the type of reasoning Renkl (1997) found associated with successful example problem solvers. In his study, Stark presented students with a series of worked examples involving probability calculation, with each example consisting of four "pages." The first page contained the givens of the problem. Each subsequent page displayed an additional step toward the solution, with the final page giving the whole solution to the problem. In order to encourage anticipative reasoning, Stark systematically inserted question marks in the examples and asked learners to attempt to anticipate what step the question mark represented. After offering an answer, learners received feedback on the accuracy of their anticipation. Stark found that learners presented with these sequentially-presented worked example outperformed their counterparts exposed to traditional worked examples on tasks that required them to transfer the solution procedures represented in the examples. One possible explanation for this learning advantages is that a formatting arrangement such as this allows the student to examine each component of the example's solution in relative isolation from the one preceding it, because learners can literally step through an example, analyzing each problem state and the transformation required to achieve the next.

Another promising line of research focusing on structural enhancements is Catrambone's (1994, 1996, 1998) work investigating the effect of labels on learning from worked examples. This work parallels research in the text-comprehension literature on the effects of signals or cues on text-processing strategies (e.g., Lorch & Lorch, 1995; Meyer & Rice, 1989). Just as organizational signals in text induce learners to change their text-processing strategy by cueing the important text content and its organizational structure, worked-example labels are intended to increase the likelihood that learners will discern the hierarchical conceptual structure of the problem contained in the example. In particular, Catrambone's research focuses on inducing learners to form subgoals, which he describes as the process of attributing a purpose or goal to a subset of steps in an example's solution that represent a meaningful conceptual piece. His research suggests that segmenting and labeling a problem so that it emphasizes discrete subgoals encourages learners to self-explain. He conjectured that the advantage arises from the learner being forced to determine the goal or function of those steps through a self-explanation process. Across a series of studies, Catrambone has convincingly demonstrated that learning can be enhanced by making the goal structure of a problem's solution explicit by using structural



manipulations, such as the use of solution step labels, or by visually isolating parts of the problem solution.

In a departure from previous work, which sought to identify techniques for limiting cognitive activities that unnecessarily burden a person's finite working memory while studying examples, Mousavi, Low, and Sweller (1995) investigated the possibility of improving example processing and problem-solving performance by incorporating the use of a dual mode of presentation. Drawing in part from Paivio's (1990) dual-coding theory, which implies the existence of separate and partially independent representational systems—one for verbal and one for visual (e.g., pictorial) information—Mousavi et al. (1995) posited that the “work on modality effects has suggested that effective working memory capacity may be enlarged by using multiple channels” (p. 319). They argue that if both systems can be used simultaneously, as opposed to overloading a single channel, the amount of information that someone is likely to retain increases. Although it is debatable whether the design of their study actually allowed them to test this supposition, Mousavi and his colleagues were able to examine whether or not an instructional device that utilized two modalities improved problem solving. To do this, they experimentally manipulated the visual and the auditory presentation of solution steps in a worked example. In one condition, students studied a geometry diagram (i.e., visual channel) while listening to an oral presentation (i.e., verbal channel) of the solution steps. Their peers in another condition studied the diagram (i.e., visual channel) and its accompanying text-based statements (i.e., visual channel).

Across a series of experiments, the authors demonstrated that learning was consistently enhanced by the dual-presentation mode. Mousavi et al. (1995) attributed this advantage to what amounts to an expansion of working memory, the result of presenting example diagrams visually and procedures aurally. In other words, increased cognitive capacity results from the information being distributed across two working memories—verbal and visual working memory—as opposed to one. Whether or not this theory is reasonable remains to be seen, but on the basis of their experimental results the authors offered specific recommendations on how to structure mixed-mode instructional materials. In particular, the authors suggested that some segments of instructional information should be presented visually while other segments should be presented aurally.

Clearly, the research reviewed in this section indicates that the structure of a worked example plays a critical role in its efficacy. Based on their respective findings, the researchers offered specific recommendations on structuring examples that will enhance a learner's ability to cognitively process an example. First, Stark's (1999) research suggests that example processing can be enhanced by sequentially presenting problem states. Catrambone's research (1994a, 1994b, 1995a, 1995b, 1996; Catrambone & Holyoak, 1990) indicates that worked examples should be structured so they emphasize conceptually-related solution steps—by visually isolating them, by labeling them, or both. With regard to presenting examples that require students to reference multiple sources of information, Mousavi and his colleagues (Mousavi et al., 1995) offer a simple solution: some segments of instructional information should be presented visually while other segments should be presented aurally.

### The Search for the Optimal Learning Environment

One advantage of using the computer as a means for delivering instruction is that it enables instructional designers to combine multiple instructional principles or components in a worked example, which may prove to substantially enhance its efficacy. For example, the results of a study by Atkinson and Derry (2000) suggests that one way to structure an example within a computer-based multimedia environment so that learning can be maximized is to create a multi-component worked example that (1) is sequential—in that it consists of a sequential presentation of problem states; (2) incorporates a second modality that is coordinated with the sequential presentation of problem states (i.e., visually presented steps coupled with verbal instructional explanations); and (3) is constructed to emphasize problem subgoals (i.e., it is subgoal-oriented). According to their study, learners exposed to these sequential, subgoal-oriented (SE/SO) examples with dual modes outperformed learners who were exposed to more traditional, simultaneous, non-subgoal-oriented examples on conceptually-based measures of problem-solving transfer. Moreover, this difference occurred despite the fact that the examples in the latter condition were also dual mode. Although the measurable effects of this type of example were not as dramatic or as pervasive as the authors predicted, it was apparent from their results that learners engaged in mathematical thinking could benefit on a conceptual level by working with sequential, subgoal-oriented, computer-based examples.

Although the examples used in the Atkinson and Derry (2000) study incorporated structural modifications that were intended to tacitly encourage learners to adopt effective self-explanation styles during example processing (i.e., anticipative reasoning due to the sequential presentation of problem states), it may be that students need explicit training in how to use this structural modification to their benefit in order to generate more dramatic learning effects. In particular, the SE/SO examples used in the Atkinson and Derry study did not contain any overt prompts (e.g., question marks in example solution) to encourage learners to engage in anticipative reasoning. Instead, the learner simply moved forward through the example and watched as problem states were successively added over a series of pages—similar to an animation, with the final page in the series representing the solution in its entirety. In the case of SE/SO examples, it may be possible to encourage learners—through more explicit prompting—to adopt good self-explanation strategies and thereby increase the effectiveness of this type of example. Stark's (1999) research suggests one technique for accomplishing this task, namely, systematically inserting blanks into the example solution.

However, although research suggests that self-explanations are a critical component in example processing and subsequent learning (e.g., Chi et al., 1989, Renkl, 1997), Renkl (in press) points out that there are a number of shortcomings associated with self-explanations. First, they may not always be correct. That is, students may derive—through the self-explanation process—an inaccurate rationale for the steps in an example's solution due to misconceptions that they may have about the procedure illustrated in the example, which may ultimately impede learning. Second, learners sometimes simply reach an impasse while studying examples from which they are unable to recover without external support. Third, learners may often leave an example having the illusion of understanding; that is, some students employ poor self-explanation strategies (e.g., passive or superficial learners) while studying an example, yet finish the task convinced that they have fully comprehended that material they have studied.

Instructional explanations, on the other hand, can help circumvent some of these disadvantages associated with self-explanations. Although they have their own set of disadvantages (e.g., not tailored to learner's prior knowledge, not generative in nature), Renkl (in press) suggests that they can be used in ways that compliment the self-explanation process. Since, according to Renkl, instructional explanations are designed by instructional designers, they are by their very nature correct. They also have the potential—if well timed—to help students overcome comprehension impasses. Finally, they can help learners discover the limits of their understanding.

Based on the review of example processing research, it appears that, ideally, a learning environment involving worked examples should not only promote successful self-explanation styles but also strive to provide accurate instructional explanations. In sum, learning environments that combine the critical components of self-explanations with those of instructional explanations may ultimately prove to be the most successful instructional format for fostering learning from worked examples. An empirical examination of this issue is certainly warranted.

The literature on computer-based tutoring systems also suggests another important issue that needs to be addressed in an effort to structure a learning environment that will optimize learning from worked examples. Although there are several computer-based tutoring systems currently under development, to date they fail to incorporate the features that are critical to successful human tutors, such as their motivational and affective features. One potential difficulty has to do with plausibility or acceptability of a computer tutor behaving in a "lifelike" manner by conveying and eliciting emotion or other behavior associated with effective human tutors (Lester & Stone, 1997). At issue is whether the same actions, statements, and gestures that human tutors use during face-to-face tutoring sessions to give encouragement, for example, will have the same effect on a student if delivered instead by a computer-based tutoring system.

Recently, one possible solution to this problem has begun to emerge in the form of animated agents with lifelike qualities that operate within computer-based learning environments. In particular, it may be possible to structure an example-based learning environment so that a life-like character can exploit verbal (e.g., instructional explanations) as well as nonverbal forms of communication (e.g., gaze, gesture) within the examples themselves in effort to promote a learner's motivation toward the task and his/her cognitive engagement in it. Lester and his colleagues (Johnson, Rickel, & Lester, in press; Lester, Stone, & Stelling, 1999; Lester, Converse, Stone, & Kahler, 1997) suggest that these life-like characters, which have been described as animated pedagogical agents, are ideally-suited to serve as coaches and guides in knowledge-based learning environments. Essentially, these animated pedagogical agents are engaging animated characters that can "couple key feedback functionalities with a strong visual presence by observing learners' progress and providing them with visually contextualized problem-solving advice" (Lester, Stone, & Stelling, 1999). These agents reside in the learning environment by appearing as animated "humanlike" characters, which allows them to exploit nonverbal communication typically reserved for human-human interactions. For example, the agent can focus a student's attention by moving around the screen, using gaze and gesture, providing nonverbal feedback, and conveying emotions. Furthermore, these agents are capable of supporting conversation-based, task-oriented dialogue by responding to either keyboard/mouse

input or—since they have a speech recognition engine—voice commands of a user. In sum, animated agents may be able to overcome the limitation of traditional “intelligent” tutors (i.e., textual dialogue only) by performing many tutorial activities in a remarkably lifelike fashion (Johnson, Rickel, & Lester, in press). But the true potential of these animated agents lies in the potential motivational benefits that may come from providing learners with a lifelike computer-based agent that is capable of producing many of the natural forms that we as humans rely on for communication, including speech, gesture, and locomotion.

Despite the potential of employing pedagogical agents to emulate the actions of human tutors while operating in a computer-based learning environment, to date very little empirical research has been conducted in this area. In fact, the literature contains only one study focused on examining the efficacy of animated pedagogical agents operating in a computer-based learning environment (Lester, Converse, Stone, Kahler, & Barlow, 1997). Although this study found that animated agents were capable of enhancing problem solving for middle school students operating in a design-centered learning environment for botanical anatomy, one study is not sufficient to establish the effectiveness of such agents.

As previously mentioned, one way to structure an example within a computer-based multimedia environment so that learning can be maximized is to create a worked example that is sequential, incorporates a second modality, and is subgoal-oriented. Are there other ways to optimize the computer-enriched delivery method of examples? This preceding discussion suggests that there are two additional structural modifications that appear worth examining, namely incorporating an optimal type of explanation (i.e., self-explanation, instructional explanations, or a combination of the two) and relying on an animated agent to deliver the optimal explanation. Perhaps an animated pedagogical agent could be incorporated into this environment to serve as a virtual tutor during problem solving, a tutor that will encourage optimal example processing. But, the question that needs to be addressed before the agent is assimilated into the learning environment is what type of interaction is the most effective? Should an agent deliver self-explanation prompts, instructional explanations, or some combination of the two types of explanations?

The purpose of the current study was to empirically examine how the effectiveness of a computer-based learning environment could be optimized through the use of an animated agent. Specifically, while holding the instructional material and the learning environment constant, three experiments were conducted in which the proposed structural enhancements were added and tested successively. The purpose of the initial experiment was to determine whether self-explanation prompts, instructional explanations, or some combination of the two was most effective as a means of enhancing proportional reasoning. The purpose of second experiment study was to examine whether the introduction of animated agent into the learning environment would enhance learning. A problem arose, however, in the second experiment that raised some additional issues that needed to be addressed by a third experiment.

### Experiment 1

The first experiment was intended to address one primary question: Are computer-based worked examples designed to encourage learners to adopt successful self-explanation styles as well as provide instructional explanations more effective at promoting learning than more conventional worked examples? Based on the available research describing the benefits of self-explanations



(e.g., Chi et al., 1989, Renkl, 1997) and Renkl's (in press) work describing the respective strengths and weaknesses of self-explanations and instructional explanations, it was predicted that students assigned to the condition that receives both self-explanation prompts and instructional explanations would outperform those in all of the other conditions. Moreover, it was predicted that the participants in the conditions that received either a self-explanation prompt or on-demand instructional explanations would outperform those in the condition lacking both features.

### Method

**Participants and Design.** One hundred undergraduate students from the educational psychology and psychology departments at Mississippi State University volunteered to participate in the study. The participants were randomly assigned in equal proportions to one of the four conditions: instructional explanation (IE), self-explanation (SE), both self-explanations and instructional explanations (SE/IE), or neither (control).

**Materials.** The pencil-and-paper materials included a demographic questionnaire, a review of proportion problems, a pretest, and a posttest. The materials used in this study consisted of instruction and problems involving multi-step mathematical proportions. These materials were adopted with permission from Atkinson and Derry (2000), which Atkinson and Derry had modified from a set originally developed by Derry, Weaver, Liou, Barker, and Salazar (1991). The questionnaire asked each student to provide information (e.g., standardized test scores, number of post-secondary mathematics courses in progress or completed) that could be used to gauge the student's level of prior knowledge. The review of the proportion reasoning described how to identify and solve proportional relationships in simple one-step word problems and included practice problems complete with answers for the students to use in checking the accuracy of their problem solutions. The pretest consisted of eleven proportion problems of varying difficulty. The pretest was designed to gauge the student's ability to solve, prior to treatment, proportion-word problems and to perform basic arithmetic operations in a variety of problem-solving contexts. The posttest consisted of five multistep word problems, in which one or more proportional situations were embedded in combination with other arithmetic operations.

**Learning Environment.** The computer-based learning environment used in this experiment was modeled after the Instructional Multimedia Modules (IMMs) designed to deliver worked-example instruction to students learning to solve proportion word problems in TiPS and in the experiment conducted by Atkinson and Derry (2000). TiPS consists of a problem-solving interface designed to help promote students' ability to model and reason about story problems (Derry, Wortham, Webb, & Jiang, 1996; Derry, Tookey, Smith, Potts, Wortham, & Michailidi, 1994; <http://www.wcer/wisc/edu/tips/>), which employs sequential, subgoal-oriented computer-based worked examples as one of its instructional components. The IMM used in this study was created by Director 6.0 (Macromedia, 1997) software—an authoring tool for multimedia productions. The IMM was modular, in that it supported a number of structural and instructional manipulations. Regardless of the instructional manipulations, the worked-example IMM contained a number of invariant structural features, including: (1) an instruction pane, for displaying the instructions for the current problem; (2) a text pane, for displaying the problem on which the worked example was based; (3) a set of voice files, each containing a segment of an expert's elaborated description of one path to solving proportion problems; (4) a control panel, allowing the user to proceed through the instructional sequence at his/her own pace; and, (5) a



workspace, for displaying the solution to the example's problem. In addition, the IMM was programmed to record all user interactions, including mouse clicks associated with use of the control panel, time spent viewing any particular segments of an example's solution, and total training time.

In all of the conditions, students were presented with an IMM containing an example word problem followed by its solution steps, with each problem state being presented sequentially. In other words, the example appeared initially unsolved. Then the learner was able to watch as every step was visually constructed over a series of frames until the final frame depicted the completely worked-out problem. The example also contained two explicit cues designed to demarcate a problem's subgoals; that is, each subgoal was visually isolated and labeled.

In particular, the IMM used in the present experiment was configurable to run in one of four modes that reflected the four conditions of the experiment. In the instructional explanation condition (IE), students were presented with an example word problem followed by its solution steps (see Figure 1). Each solution step included instructional elaborations designed to highlight what was occurring in that step (e.g., "First, we need to set up a proportional relationship to determine the production rate"). For the self-explanation condition (SE), rather than being presented with the solution steps, the students were asked to enter the answer to just the first part of the problem, after which they viewed the correct step (see Figure 2). Then, they were asked about the next step and the process repeated until the example was complete. The intent was to give the student the opportunity to anticipate the solution to the next step in the example and then to judge the accuracy of his/her answer by comparing it to the one provided by the computer. The self-explanation/ instructional explanation condition (SE/IE) had the qualities of both the other conditions. First, the student was asked for input about the step. Then, when viewing the correct step, the elaborated instructions appeared as in the first condition. Finally, in the control condition, the IMM showed only the solution steps for the example problem without elaboration or opportunity for input.

In every condition, each example problem was followed by a practice problem presented on the computer that was parallel in structure to the example. The student was required to enter a response to the practice problem before he or she was given the final answer to the problem. The answers to the practice problems did not include solutions to problem steps nor any explanation about the solution. The four examples and their practice problems also contained periodic questions checking the student's perceptions of understanding and difficulty related to each problem. After viewing the example students were asked if they had understood the problem and they were asked to rate the difficulty of the problem based on a five-point Likert scale that ranged from very easy to very difficult. After each practice problem they were asked if they understood the example better after having worked a practice problem and they were again asked to rate the difficulty of the example problem using the aforementioned scale.

The amounts of time spent on each example, practice problem, or posttest problem were recorded in seconds from the time the computer showed the first part of the problem to the time the participant entered his or her answer. In order to control for amount of time spent viewing examples, the steps in the conditions with less information displayed appeared after a slight delay. In this manner, each of the conditions spent approximately the same amount of time

viewing each worked example. The answers that were entered for each step in the SE and SE/IE conditions, and for the practice problems and posttest problems for all conditions were similarly recorded.

Procedure. The students participated in the experiment in two sessions that were not more than a week apart. During both session, the participants worked independently and were tested individually. For the majority of the participants, the first session lasted approximately forty minutes. During the first session, the participants filled out a demographic questionnaire, then read through an eight-page review on solving proportion problems. The review also contained three basic problems that the participants were encouraged to try, followed by their complete solutions. When the subjects completed their review, they were given a pretest.

The second session of the experiment took place in one of two computer labs and its structure varied according to condition. During this session, participants were asked to study the four sets of items provided by the IMM, with each set consisting of a condition-specific worked example presented on the computer monitor along with a paired isomorphic practice problem presented on the monitor but solved on accompanying paper. Immediately following the instructional portion of the experiment, the participants concluded with a posttest presented on the computer lasting approximately one hour. The test questions were presented individually on the computer screen, but the students were asked to show their work on a separate paper packet. They were required to enter an answer before they could proceed to the next problem and they could not return to any of the preceding problems.

Scoring. The protocols generated during problem-solving practice and on the posttest were coded for conceptual scores. That is, each item was awarded a complex score, ranging from 0 to 3, depending upon the degree to which the participant's solution was conceptually accurate (e.g., 0 = no evidence of the student's understanding the problem, 3 = perfect understanding of the problem). One experimenter "blind" to the condition independently coded each protocol. A second experimenter coded a random set of twenty protocols. A Pearson correlation coefficient was calculated to assess interrater reliability. The tests were coded with a reliability of .92 across the practice and posttest problems. Discussion and common consent was used to resolve any disagreement between coders.

## Results

Analysis of Covariance (ANOVA) was used to examine for differences across the four conditions, on a number of measures, including: (1) reported example understanding before exposure to an paired practice problem; (2) reported example understanding after exposure to an paired practice problem (3) reported example difficulty before exposure to an paired practice problem; (4) reported example difficulty after exposure to an paired practice problem; (5) performance on practice problems; (6) performance on posttest; (7) total instructional time; and (8) time on posttest. Also, the SE and SE/IE conditions were compared on correctness of anticipations. For statistical control purposes, the data were adjusted for one covariate, namely, the pretest. In the event the conditions main effect was found to be significant, pairwise comparisons among the various treatment conditions were examined using Fisher-Hayter's procedure, based on a familywise  $\alpha$  of .05. The Fisher-Hayter procedure was chosen because of its superior power compared to other pairwise comparison procedures under similar

circumstances (Seaman, Levin, & Serlin, 1991). Adjusted means for ANCOVA's that were found to be significant are reported in Table 1.

For reported example understanding before practice, an ANCOVA revealed that the means of four treatments differed statistically from one another,  $F(3, 95) = 4.06$ ,  $MSE = .48$ ,  $p < .01$ ,  $\eta^2 = .11$ . Supplemental Fisher-Hayter comparisons revealed that the participants in the IE condition reported understanding the examples better than the participants in the SE and SE/IE conditions. However, a separate ANCOVA of conditions-related difference on reported example understanding after practice revealed no significant effect,  $F(3, 95) = .33$ ,  $MSE = 2.30$ ,  $p > .05$ . Also, there was no significant conditions effect on the reported example difficulty, either before practice,  $F(3, 95) = .56$ ,  $MSE = 10.78$ ,  $p > .05$ , or after practice,  $F(3, 95) = 1.32$ ,  $MSE = 9.38$ ,  $p > .05$ .

An ANCOVA conducted on the measure consisting of the performance on practice problems indicated that there was an overall difference among the conditions,  $F(3, 94) = 3.73$ ,  $MSE = 8.87$ ,  $p < .05$ ,  $\eta^2 = .11$ . Fisher-Hayter comparison of this measure revealed that the participants in the IE condition were statistically superior to their peers in the SE condition in terms of the average performance on practice problems.

A significant conditions main effect,  $F(3, 95) = 9.01$ ,  $MSE = 7.93$ ,  $p < .001$ ,  $\eta^2 = .22$ , was also found on the performance on posttest. Fisher-Hayter comparisons showed that the participants in the SE condition performed more poorly than those in each of the other groups, namely the control, IE, and SE/IE. Furthermore, analysis of the group means revealed that IE condition was statistically superior to the control and SE/IE in terms of performance on posttest.

An ANCOVA performed on the total instructional time indicated that the four conditions differed significantly in their mean on this measure,  $F(3, 95) = 11.54$ ,  $MSE = 117.51$ ,  $p < .001$ ,  $\eta^2 = .27$ . Supplemental Fisher-Hayter comparisons revealed that participants in the control and IE condition spend less time in instruction than their counterparts in the SE or SE/IE conditions. For the time on posttest, there was also a significant conditions main effect,  $F(3, 95) = 4.33$ ,  $MSE = 45.99$ ,  $p < .001$ ,  $\eta^2 = .12$ . Fisher-Hayter comparisons of the time on posttest showed that IE participants spent significantly more time on the posttest than their peers in the SE and SE/IE conditions.

An ANCOVA performed on the correctness of anticipations indicated that SE and SE/IE conditions differed significantly in their mean on this measure,  $F(1, 47) = 11.81$ ,  $MSE = .02$ ,  $p < .001$ ,  $\eta^2 = .20$ . Participants in the SE/IE condition ( $M = .40$ ) provided more accurate anticipations of the next step than their peers assigned to the SE condition ( $M = .26$ ).

### Discussion.

In sum, participants in the IE condition reported understanding the examples better than the participants in the SE and SE/IE conditions (medium-to-large effect size) before practice, although this difference did not appear on the sum of reported example understanding after practice. Despite the fact that participants in the IE condition spent less time engaged in instruction than their counterparts in the SE condition, the participants in the IE condition outperformed their peers in the SE condition in terms of conceptual performance on practice

problems (medium-to-large effect size) and posttest (large effect size). In fact, participants in the SE condition performed more poorly than those in each of the other groups, namely the control and SE/IE (large effect size).

A final noteworthy finding was that the participants in the SE/IE condition provided more accurate anticipations of the next step than their peers assigned to the SE condition (large effect size). As previously mentioned, for the self-explanation condition (SE), rather than being presented with the solution steps, the students were asked to enter the answer to just the first part of the problem, after which they viewed the correct step. Then, they were asked about the next step and the process repeated until the example was complete. The intent was to give students an opportunity to anticipate the solution to the next step in the example and then to judge the accuracy of their answers by comparing their to the one provided by the computer. According to the results of this experiment, the participants in the SE/IE condition provided more accurate anticipations of the next step than their peers assigned to the SE condition. Thus, instructional explanations led to more accurate anticipations.

Although it seemed reasonable at the outset of this experiment to expect that better learning outcomes could be achieved by creating a dual component worked example that combines periodic self-explanation prompts with instructional explanations, this study did not find support for such a combination. In fact, the results of this study suggest that worked examples containing instructional explanations only (IE) are superior to examples containing self-explanation prompts only (SE). Overall, participants assigned to the IE condition reported understanding the examples more often and had more conceptually accurate solutions to the practice and posttest problems than their SE peers despite the fact that they spent less time on the instructional segment of the experiment. In sum, it is apparent from the evidence compiled in the present study that students engaged in mathematical thinking can benefit on a conceptual level by working in a learning environment that provides access to instructional explanations of each example step without concern for explicitly promoting the self-explanation process.

## Experiment 2

Having established that instructional explanations were more effective than self-explanation prompts in my learning environment, Experiment 2 was designed to extend this finding by introducing animated pedagogical agents into the learning environment. Specifically, since instructional explanations were found to be effective in Experiment 1, Experiment 2 examined whether an animated agent programmed to deliver instructional explanations was more capable of supporting learning in the context of example-based instruction than conventional single mode examples (e.g., text-only or voice-only examples). Thus, Experiment 2 addressed the following new questions: (1) Are worked examples, when coupled with animated pedagogical agents, more effective at promoting learning than more conventional single-mode worked examples (predicted to be the case); and, (2) are these animated pedagogical agents more capable of promoting learning than dual-mode worked examples that include verbal self-explanation prompts and verbal on-demand instructional explanations but simply lack the interactive, life-like characters examples (predicted to hold true)?

## Method

Participants and Design. Thirty undergraduate students from the educational psychology and psychology departments at a State University of New York—College at Oneonta volunteered to participate in the study. The participants were randomly assigned in equal proportions to one of the three conditions: voice + agent, voice-only, or text-only control.

Materials. All of the pencil-and-paper materials used in this experiment were the same as the materials employed in the first experiment. Additionally, this experiment incorporated a brief (five-item) questionnaire or Likert-type affective scale where participants were asked to judge the effectiveness of the instructional program on a seven-point scale. Specifically, the participants were asked to rate the instruction program that they had completed on interest (from boring to interesting) and comprehension (from very confusing to easily understood). They were also asked to indicate their agreement or disagreement with the following statements: "this is a good way to learn," "I would like instruction in my courses like this," and "I had trouble focusing my attention during this instruction." The participants were instructed to respond by answering from strongly agree to strongly disagree.

Learning Environment. The IMM used in Experiment 1 was modified for the purposes of the second experiment. Across all of the conditions in this experiment, each sequentially-presented solution step (i.e., subgoal) included instructional elaborations designed to highlight what was occurring in that step (e.g., "First, we need to set up a proportional relationship to determine the production rate"). In the voice + agent condition, the instructional elaborations were aurally delivered by Peedy, an agent created by an off-the-shelf piece of software capable of producing animated agents known as Microsoft Agent (see Figure 3). According to Microsoft:

Microsoft® Agent is a set of programmable software services that supports the presentation of interactive animated characters within the Microsoft Windows® interface. Developers can use characters as interactive assistants to introduce, guide, entertain, or otherwise enhance their Web pages or applications in addition to the conventional use of windows, menus, and controls. Microsoft Agent enables software developers and Web authors to incorporate a new form of user interaction, known as conversational interfaces, that leverages natural aspects of human social communication. In addition to mouse and keyboard input, Microsoft Agent includes optional support for speech recognition so applications can respond to voice commands. Characters can respond using synthesized speech, recorded audio, or text in a cartoon word balloon. The conversational interface approach facilitated by the Microsoft Agent services does not replace conventional graphical user interface (GUI) design. Instead, character interaction can be easily blended with the conventional interface components such as windows, menus, and controls to extend and enhance your application's interface. Microsoft Agent's programming interfaces make it easy to animate a character to respond to user input. Animated characters appear in their own window, providing maximum flexibility for where they can be displayed on the screen. (<http://msdn.microsoft.com/workshop/c-frame.htm#/workshop/imedia/agent>)

The voice was created from a text-to-speech engine provided by Microsoft on their Agent website. For the voice-only condition, student listened to a tutor's voice (the same computer-generated voice used in the voice + agent condition) deliver the instructional explanations used to highlight the solution steps without an agent present. In the text-only or control condition, students read the textual explanations without an agent being present (see Figure 1).



Procedure. Similar to Experiment 1, the students participated in the study in two sessions that were not more than a week apart. The first session lasted approximately forty minutes, during which time the participants filled out a demographic questionnaire, then read through an eight-page review on solving proportion problems, and solved the pretest. The second session of the study was conducted in a learning center equipped with eight individual computer workstations. In this session, participants were treated in groups ranging in size from 2 to 12 participants. Each participant was seated at a separate computer workstation, provided headphones, and given specific instructions not to converse with the other participants. Participants worked for approximately ninety minutes, during which time they completed a set of computer-based material, including instruction, a posttest, and an affective scale.

Scoring. As in Experiment 1, the protocols generated during problem-solving practice and on the posttest were coded for conceptual scores. The tests were coded with an inter-rated reliability of .97 across the practice and posttest problems.

### Results

To examine for differences across the three conditions, performance was examined across the same measures as Experiment 1 in addition to the new measure introduced to this experiment, namely the instructional program evaluation questionnaire. For statistical control purposes, the data on the performance measures were adjusted for one covariate, namely, the pretest. Any significant differences were followed up with Fisher's LSD, based on a familywise  $\alpha$  of .05.

An ANCOVA conducted on the measure consisting of the performance on practice problems indicated that there was an overall difference among the conditions,  $F(2, 26) = 4.17$ ,  $MSE = 6.90$ ,  $p < .05$ ,  $\eta^2 = .24$ . Fisher's LSD comparison of this measure revealed that the participants in the text-only control condition were statistically superior to their counterparts in the voice-only condition in terms of the average performance on practice problems (see Table 2).

Analysis of the questionnaire items revealed several significant differences. An ANCOVA conducted on the questionnaire item that gauged interest in the material indicated that there was an overall difference among the conditions,  $F(2, 26) = 3.29$ ,  $MSE = 1.11$ ,  $p = .05$ ,  $\eta^2 = .20$ . Fisher's LSD comparison revealed that the participants in the text-only condition reported higher interest than their peers in the voice-only condition. An ANCOVA conducted on the questionnaire item that gauged comprehension of the material indicated that there was an overall difference among the conditions,  $F(2, 26) = 4.112$ ,  $MSE = 1.76$ ,  $p < .05$ ,  $\eta^2 = .24$ . According to Fisher's LSD analysis of this measure, participants in the text-only and agent conditions reported understanding the material (i.e., comprehension) better than their counterparts in the voice-only condition. An ANCOVA conducted on the questionnaire item that asked whether this is a good way to learn indicated that there was an overall difference among the conditions,  $F(2, 26) = 3.81$ ,  $MSE = 1.33$ ,  $p < .05$ ,  $\eta^2 = .23$ . The results from the post hoc Fisher's LSD analysis indicated that the participants in the text-only condition reported more favorable responses to the learning capacity of the instruction than the participants assigned to the voice-only condition.

### Discussion.

In sum, the participants in the text-only control condition were statistically superior to their counterparts in the voice-only condition on a variety of measures (all amounted to large effect sizes), including: average sum of conceptual scores on practice problems, interest in the material, comprehension of the material, and preference for the instruction as a good way to learn. The only difference found between the participant in the agent condition and the other two conditions appeared on the comprehension question, where participants in the text-only and agent conditions reported understanding the material (i.e., comprehension) better than their counterparts in the voice-only condition. Unexpectedly, the findings of Experiment 2 also contradicted the literature that suggests that dual-mode presentation techniques (e.g., auditory text and visual diagrams) are typically superior to conventional, single-modality formats (e.g., visual text and visual diagrams) or visual-only formats (Tindall-Ford, Chandler, & Sweller, 1997).

As result of informal exit interviews, it became apparent that the outcome of Experiment 2 could potentially be attributed to the text-to-speech engine used in the two conditions that relied on aurally presented instructional explanations, namely the voice + agent and voice-only conditions. In fact, it appeared to compromise performance. This outcome may be the result of limitations associated with the text-to-speech engine's voice quality, which Microsoft readily acknowledges can sound synthesized:

Most text-to-speech engines can render individual words successfully. However, as soon as the engine speaks a sentence, it is easy to identify the voice as synthesized because it lacks human prosody -- i.e., the inflection, accent, and timing of speech. For this reason, most text-to-speech voices are difficult to listen to and require concentration to understand, especially for more than a few words at a time (<http://www.microsoft.com/IIT/OnlineDocs/intro2tts.html>).

Microsoft also suggests that the "[t]ext-to-speech should be used to audibly communicate information to the user, when digital audio recordings are inadequate [and that] generally, text-to-speech is better than audio recordings when: (1) audio recordings are too large to store on disk or expensive to record; and (2) audio recording is impossible because the application doesn't know ahead of time what it will speak" (<http://www.microsoft.com/IIT/OnlineDocs/intro2tts.html>).

### Experiment 3

As a result of the unexpected findings and the informal feedback from the participants in Experiment 2, Experiment 3 was conducted to replicate and extend the conditions of the earlier experiment. For Experiment 3, the text-to-speech engine used in the voice + agent and voice-only conditions in Experiment 2 was replaced with audio recordings of a human tutor and the issues raised in that experiment were reexamined. Additionally, the third experiment was designed, in part, to test for a modality effect. Thus, this experiment was designed to address the following questions: (1) Are worked examples, when coupled with animated pedagogical agents, more effective at promoting learning than more conventional single-mode worked examples; (2) are these animated pedagogical agents more capable of promoting learning in comparison to dual-mode worked examples that include verbal self-explanation prompts and verbal on-demand instructional explanations but simply lack the interactive, life-like characters examples; and (3) are two sensory modes superior to one (i.e., is there a modality effect)?

## Method

**Participants and Design.** Fifty undergraduate students from several educational psychology and psychology courses at a large university located in the south volunteered to participate in the study. The experiment consisted of a 2 x 2 factorial design with a control. The first factor was the modality of explanation (voice or text); the second factor was the presence or absence of an animated agent (agent no agent). Thus, the participants were randomly assigned in equal proportions to one of the five conditions: voice + agent, text + agent, voice only, text-only, or control, which did not contain any instructional explanations or an agent.

**Materials.** All of the materials were the same as those used in the second experiment, with the exception of the posttest. The posttest was modified to include a total of six items. Three of the items were near transfer problems, which were similar in structure to an example/practice problem pair. The other three problems on the posttest could be considered far transfer problems that required knowledge of proportions but were not explicitly similar in format to any of the example/practice problem pairs. In particular, these multi-step word problems varied from the example/practice problem pairs along a number of dimensions, including story and structural context (i.e., proportional situation embedded among different arithmetic operations), as well as number of proportion situations included.

**Learning Environment.** In all of the conditions, students were presented with an IMM containing an example word problem followed by its solution steps, with each problem state being presented sequentially. In other words, the example appeared initially unsolved. Then the learner was able to watch as every step was visually constructed over a series of frames until the final frame depicted the completely worked-out problem. The example also contained two explicit cues designed to demarcate a problem's subgoals; that is, each subgoal was visually isolated and labeled. For the treatment conditions (excluding the control), each solution step or subgoal included the instructional elaborations from Experiment 1 designed to highlight what was occurring in that step (e.g., "First, we need to set up a proportional relationship to determine the production rate"). Since the animation services provided by Microsoft Agent permit audio files to be used for character's spoken output, Peedy was programmed in the voice + agent condition to deliver the recoded audio explanations created by a human tutor (see Figure 3). The software also automatically synchronized Peedy's mouth to the human tutor's voice by using the characteristics of the audio file. For the text + agent condition, rather than being presented with the instructional elaborations aurally, the students were presented with textual instructional explanations (see Figure 4). In the voice-only condition, student listened to the human tutor's voice (the same voice used in the voice + agent condition) highlighting solution steps but no agent was present. In the text-only condition, students read the textual explanations without an agent being present (see Figure 1). Finally, in the control condition, only the solution steps for the example problem were presented without instructional elaborations or agent.

**Procedure.** Similar to Experiments 1 and 2, the students participated in the study in two sessions that were not more than a week apart. The first session was identical to the initial session of Experiments 1 and 2. The second session of the study was conducted in a learning center equipped with eight individual computer workstations. In this session, participants were treated in groups ranging in size from 4 to 8 participants. Each participant was seated at a separate computer workstation, provided headphones, and given specific instructions not to

converse with the other participants. Participants worked for approximately ninety minutes, during which time they completed a set of computer-based material, including an instruction, a posttest, and an affective scale. Immediately following the instructional portion of the study, the participants concluded with a six-item posttest (three near transfer items and three far transfer items) presented on the computer.

Scoring. Comparable to Experiments 1 and 2, the protocols generated during problem-solving practice and on the posttest were coded for conceptual scores. The tests were coded with an inter-rater reliability of .99 across the practice and posttest problems.

### Results

The analysis consisted of two complimentary  $\alpha$ -controlled sets of analyses on each dependent measure—the same measures as Experiment 2. First, comparisons between each of the “treatment” conditions and the control condition were made using Dunnett’s multiple comparison procedure based on a familywise  $\alpha$  of .05. For statistical control purposes, the data were adjusted for one covariate, namely, the pretest. Then, a  $2 \times 2$  (agent  $\times$  voice) analysis of covariance (ANCOVA) was conducted using the pretest as a covariate, also based on a familywise  $\alpha$  of .05

The major results of the comparisons between each of the “treatment” conditions and the control condition are as follows: (1) voice + agent participants were statistically superior to the control participants on measures of both near and far transfer, (2) voice + agent participants and voice-only participants were statistically superior to the control participants reported example understanding before and after exposure to practice problems, and (3) voice + agent participants reported significantly lower levels of perceived example difficulty before exposure to practice problems than their control counterparts (see Table 3).

On the measure of reported example understanding before practice, an ANCOVA revealed a main effect for voice,  $F(1, 35) = 5.01$ ,  $MSE = .29$ ,  $p < .05$ ,  $\eta^2 = .13$ . The participants in the voice conditions reported understanding the examples better than the participants in the no-voice conditions. There were no other main effect or interaction. A separate ANCOVA on reported example understanding after practice revealed a significant main effect for voice,  $F(1, 35) = 4.00$ ,  $MSE = 1.08$ ,  $p = .05$ ,  $\eta^2 = .10$ . Like the results from the reported example understanding before practice, participants in the voice conditions reported understanding the examples better than the participants in the no-voice conditions after exposure to practice problems. Also, there was a significant main effect for voice on reported example difficulty, before practice,  $F(1, 35) = 6.42$ ,  $MSE = 10.66$ ,  $p < .05$ ,  $\eta^2 = .16$ . Similar to the results from the reported example understanding before and after practice, participants in the voice conditions reported that the examples were less difficult than the participants in the no-voice conditions before exposure to practice problems.

For near transfer, an ANCOVA revealed a significant main effect for voice,  $F(1, 35) = 5.91$ ,  $MSE = 3.05$ ,  $p < .05$ ,  $\eta^2 = .15$ . The participants who listened to the explanations outperformed their peers who had to read the explanations on near transfer. Neither the main effect for agent nor its interaction with voice was significant on this measure. Although overall far transfer measure was not significant, a closer examination revealed a significant interaction between

agent and voice on one item,  $F(1, 35) = 4.62$ ,  $MSE = .71$ ,  $p < .05$ ,  $\eta^2 = .12$ . Examination of the mean scores suggest a disordinal interaction, that is, the effects of the agent factor reverse themselves as the levels of the voice factor change. Specifically, for the participants provided with an agent, the voice group obtained a higher score than the no-voice (text-only) group. For participants not provided with the agent, the no-voice (text-only) group obtained a higher score than the voice group.

Finally, an analysis of the five questionnaire items revealed no significant differences on any of the items,  $F < 1.6$ .

### Discussion.

The results from the first set of analyses suggest that participants in the voice + agent, condition outperformed their control counterparts on measures of both near and far transfer, reported example understanding before and after exposure to practice problems, as well as perceived example difficulty before exposure to practice problems. Moreover, voice-only participants were statistically superior to the control participants who reported example understanding before and after exposure to practice problems.

According to the second set of analyses, the participants in the voice conditions reported understanding the examples better than the participants in the no-voice conditions before practice (large effect size) as well as after exposure to practice problems (medium-to-large effect size). Similarly, the participants in the voice conditions reported that the examples were less difficult than the participants in the no-voice conditions before exposure to practice problems (large effect size). Finally, the participants who were exposed to the explanations aurally (i.e., voice conditions) outperformed their peers who were exposed to the text-based explanations on the measure of near transfer on the posttest (large effect size).

**Significance:** The present study examined ways in which to structure computer-based worked examples to optimize their effectiveness. The results of Experiment 1 suggest that an example-based learning environment that combines the critical components of self-explanations with those of instructional explanations is not the most successful environment for fostering learning from worked examples. In fact, worked examples containing instructional explanations alone appear to encourage learners to produce more conceptually accurate solutions during concurrent as well as subsequent problem solving. The results of experiment 2 also suggest that worked examples containing instructional explanations are superior to examples containing self-explanation prompts.

Experiment 2 suggests that the text-to-speech engine that Microsoft provides in conjunction with the agent software may be detrimental to learning. As evidence, the participants in the single-mode, text-only control condition were statistically superior to their counterparts in the dual-mode, voice-only condition on conceptually accurate practice problems solution. In addition, learners exposed to the text-only control reported higher interest in the material, higher comprehension of the material, and higher preference for the instruction as a good way to learn than their counterparts in the voice-only condition. The only indication that animated pedagogical agents may be more capable of promoting learning in comparison to conventional single-mode worked examples came from the questionnaire where participants in the agent



condition as well as the text-only condition reported understanding the material (i.e., comprehension) better than their counterparts in the voice-only condition.

Experiment 3 establishes modest support for the claim that animated pedagogical agents support learning from worked examples. Although the measurable effects of this type of example were not as dramatic or as pervasive as might be expected, it is apparent from the evidence compiled in the present study that students engaged in mathematical thinking can benefit on a variety of levels (e.g., cognitive, affective) by working within a learning environment that contains an animated pedagogical agent, in particular, an agent capable of delivering instruction aurally and using forms of nonverbal communication to support learning. Along those lines, this study also provides additional empirical support for the proposal that example processing and problem-solving performance can be improved by incorporating the use of a dual mode of presentation in example-based instruction (see Mousavi, Low, & Sweller, 1995).

More generally, the use of computers in this multi-experiment study to present worked examples demonstrates that computer-based multimedia environments offer an effective medium in which example-oriented instruction can be supported. In particular, this study suggests that instructional prescriptions derived from research in the context of print-delivered media (e.g., subgoal labeling) can be successfully transferred to computer-delivered media. Furthermore, this study demonstrates that this type of medium affords instructional designers an environment in which independent instructional prescriptions regarding example-based instruction can be combined so that they function simultaneously. In contrast to a book-based medium, computers also provide a more favorable environment in which to implement some forms of effective instruction, such as the coordination of the visual presentation of sequential problem states with an auditory description of each of those states (Mayer, 1997; Mousavi et al., 1995). Finally, computer-based instruction has the potential to increase learners' conceptual understanding in mathematics by creating an environment ideally suited for permitting them to observe variance over time, a critical prerequisite to recognizing invariance in mathematics (Kaput, 1992).

Finally, the results of this study suggest a number of avenues for future research, some designed to address shortcomings of the present study, as well as others that direct attention to several productive research questions raised in the present study. Foremost among these recommendations is the need to replicate the findings discussed in the preceding pages. In particular, the findings from Experiment 3 need to be empirically replicated in order to strengthen the claim that animated pedagogical agents support learning in example-based learning environments. Future research should also attempt to empirically establish differences between the effectiveness of text-to-speech engines versus audio recordings of human tutors. Finally, while these findings might generalize to other areas of mathematics, research should examine whether these findings generalize across other semantically-rich domains.

## References

- Atkinson, R. K., & Derry, S. J. (2000). Computer-based examples designed to encourage optimal example processing: A study examining the impact of sequentially presented, subgoal-oriented worked examples. In B. Fishman & S. F. O'Connor-Divelbiss (Eds.), Proceedings of the Fourth International Conference of Learning Sciences (pp. 132-133). Hillsdale, NJ: Erlbaum.
- Catrambone, R. (1994). Improving examples to improve transfer to novel problems. Memory & Cognition, 22, 606-615.
- Catrambone, R. (1996). Generalizing solution procedures learned from examples. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1020-1031.
- Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. Journal of Experimental Psychology: General, 127, 355-376.
- Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13, 145-182.
- Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. Journal of Educational Psychology, 79, 347-362.
- Derry, S. J., Tookey, K., Smith, C., Potts, M. K., Wortham, D. W., & Michailidi, A. (1994). Psychological foundations of the TiPS system: A handbook for system 1.0 (Technical Report). Madison, WI: Wisconsin Center for Education Research.
- Derry, S., Weaver, G., Liou, Y., Barker, J., & Salazar, E. (1991). Inducing flexible transfer in novice problem solvers: Effects of three instructional approaches. Unpublished manuscript.
- Derry, S. J., Wortham, D., Webb, D., & Jiang, N. (1996). Tutorials in problem solving (TiPS): Toward a schema-based approach to instructional tool design. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.
- Johnson, W. L., Rickel, J. W., & Lester, J. C. (in press). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. International Journal of Artificial Intelligence in Education.
- Kaput, J. J. (1992). Technology and mathematics education. In D. A. Grouws (Ed.), Handbook of research on mathematics and learning: A project of the National Council of Teachers in Mathematics (pp. 515-556). New York: Macmillan.
- LeFevre, J., & Dixon, P. (1986). Do written instructions need examples? Cognition and Instruction, 3, 1-30.
- Lester, J. C., Converse, S. A., Stone, B. A., Kahler, S. E., & Barlow, S. T. (1997). Animated pedagogical agents and problem-solving effectiveness. In B. du Boulay & R. Mizoguchi (Eds.), Proceedings of the Eighth World Conference on Artificial Intelligence in Education (pp. 23-30). Washington, DC: IOS Press.
- Lester, J. C., Converse, S., Stone, B., Kahler, S., & Barlow, T. (1997). Animated pedagogical agents and problem-solving effectiveness: A large-scale empirical evaluation. Proceedings of the Eighth World Conference on Artificial Intelligence in Education (pp. 23-30). Kobe, Japan.
- Lester, J. C., Stone, B. A., & Stelling, G. D. (1999). Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. User Modeling and User-Adapted Interaction, 9, 1-44.
- Lorch, R. F., & Lorch, E. P. (1995). Effects of organizational signals on text-processing strategies. Journal of Educational Psychology, 87, 537-544.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? Educational Psychologist, 32, 1-19.

Mayer, R. E., Sims, V., & Tajika, H. (1995). A comparison of how textbooks teach mathematical problem solving in Japan and the United States. American Educational Research Journal, 32, 443-460.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. Journal of Educational Psychology, 87, 319-334.

Paas, F., & Van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. Journal of Educational Psychology, 86, 122-133.

Paivio, A. (1990). Mental representations: A dual coding approach. New York: Oxford University Press.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. Cognitive Science, 21, 1-29.

Renkl, A. (in press). Learning mathematics from worked-out examples: Analyzing and fostering self-explanations. European Journal of Psychology in Education.

Seaman, M. A., Levin, J. R., & Serlin, R. C. (1991). New directions in pairwise comparisons: Some powerful and practicable procedures. Psychological Bulletin, 110, 577-586.

Stark, R. (1999). Lernen mit Lösungsbeispielen. Der Einfluß unvollständiger Lösungsschritte auf Beispielelaboration, Motivation und Lernerfolg [Learning by worked-out examples. The impact of incomplete solution steps on example elaboration, motivation, and learning outcomes]. Bern, CH: Huber.

Sweller, J. (1993). Some cognitive processes and their consequences for the organisation and presentation of information. Australian Journal of Psychology, 45, 1-8.

Tindall-Ford, Chandler, & Sweller, J. (1997). When two sensory modes are better than one. Journal of Experimental Psychology: Applied, 3, 257-287.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. Cognition and Instruction, 7, 1-39.

Table 1.

Adjusted (For Pretest) Mean Performance on the Significant Outcome Measure of Experiment 1

Measure	Condition			
	Control	IE	SE	SE/IE
Reported Example Understanding Before Practice	3.82	3.92	3.38	3.40
Performance on Practice Problems	9.59	10.45	7.69	9.61
Performance on Posttest	8.16	10.18	6.01	7.94
Instructional Time	32.51	34.98	45.47	47.43
Time on Posttest	20.47	22.53	18.40	15.92

Table 2.

Adjusted (For Pretest) Mean Performance on the Significant Outcome Measure of Experiment 2

Measure	Condition		
	Text Only	Voice Only	Voice + Agent
Performance on Practice Problems	11.90	8.52	10.70
Interest	4.90	3.77	4.74
Comprehension	5.20	3.67	5.13
Learn	5.80	4.40	5.03



Table 3.

Adjusted (For Pretest) Mean Performance on the Significant Outcome Measure of Experiment 3

Measure	Condition				
	Control	Text Only	Voice Only	Text + Agent	Voice + Agent
Reported Example Understanding Before Practice	3.05	3.70	3.97	3.50	4.00
Reported Example Difficulty Before Practice	13.10	11.55	10.60	12.10	7.75
Performance on Practice Problems	7.42	9.43	10.13	9.00	9.53
Reported Example Understanding After Practice	2.01	3.09	3.58	2.50	3.39
Posttest – Near Transfer	5.35	5.89	6.99	5.60	7.18
Posttest – Far Transfer	3.81	4.71	5.18	4.30	6.11
Instructional Time	34.35	36.37	34.61	37.60	34.27
Time on Posttest	28.46	26.19	28.95	21.90	28.39

### Figure Captions

Figure 1. Instructional explanation (IE) condition from Experiments 1 and 2.

Figure 2. Self-explanation (SE) condition from Experiment 1.

Figure 3. Voice + agent condition from Experiments 2 and 3.

Figure 4. Text + agent condition from Experiments 3.

Instructions

Partial Amount 1

$\$27.50 \times 25 = \$687.50$

Partial Amount 1

$\$24.75 \times 11 = \$272.25$

Total Amount 1

$\$687.50 + \$272.25 = \$959.75$

Amount 2

----- = -----

Problem Text

Susan is planning a reception dinner for 36 people and is considering two restaurants. Restaurant A has a special banquet menu which they serve at a scaled rate of \$27.50 per person for the first 25 people, then \$24.75 per person for each additional person. Restaurant B does not have a banquet menu or a scaled rate. However, Susan knows that her friend John hosted a reception dinner for 22 people there and it cost him \$539.50. If Susan uses what John spent on 22 people to estimate her cost for 36 people at Restaurant B, which restaurant would she conclude is more expensive, Restaurant A or Restaurant B? And by how much?

Control Panel

⏮

⏪

⏩

⏭

Rev Back Play FF Exit

Quit

7

8

9

÷

4

5

6

×

1

2

3

+

0

.

%

=

Next we need to set up a proportional relationship to determine the cost at Restaurant B.

26

### Instructions

### Problem Text

Susan is planning a reception dinner for 36 people and is considering two restaurants. Restaurant A has a special banquet menu which they serve at a scaled rate of \$27.50 per person for the first 25 people, then \$24.75 per person for each additional person. Restaurant B does not have a banquet menu or a scaled rate. However, Susan knows that her friend John hosted a reception dinner for 22 people there and it cost him \$539.50. If Susan uses what John spent on 22 people to estimate her cost for 36 people at Restaurant B, which restaurant would she conclude is more expensive, Restaurant A or Restaurant B? And by how much?

### Control Panel

Rev Back Play FF Exit

Partial Amount 1  
 $\$27.50 \times 25 = \$687.50$

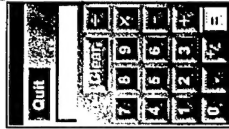
Partial Amount 1  
 $\$24.75 \times 11 = \$272.25$

Total Amount 1  
 $\$687.50 + \$272.25 = \$959.75$

Amount 2

?

Answer



### Instructions

Initial Amount

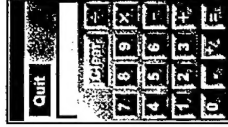
4 People 55 People

----- = ----- X

\$1,377.97

$4X = 1377.97 * 55$

$X = (1,377.97 * 55) / 4 = 18,947.09$



### Problem Text

A local travel agent, who is offering a special package for groups of students interested in taking a spring break trip to Mexico, has recruited John, a senior, to lead a group. John is told that a group of 4 can purchase a vacation package, including airfare and accommodations, for \$1,377.97. The travel agent has also offered an additional 15% discount for groups of 40 or more. John's group has 55 people. As a group, how much do they have to pay?

### Control Panel



Instructions

A local travel agent, who is offering a special package for groups of students interested in taking a spring break trip to Mexico, has recruited John, a senior, to lead a group. John is told that a group of 4 can purchase a vacation package, including airfare and accommodations, for \$1,377.97. The travel agent has also offered an additional 15% discount for groups of 40 or more. John's group has 55 people. As a group, how much do they have to pay?

Control Panel

⏮

⏪

⏩

⏭

⏹

Rev

Back

Play

FF

Exit

Initial Amount

4 People

\$1,377.97

=


55 People

X

4X = 1377.97 \* 55

X = (1,377.97 \* 55) / 4 = 18,947.09

So, the travel package for John's group will cost \$18,947.09.



Quit

Top

7

8

9

4

5

6

1

2

3

0

.

/

=

29



**Publications, Abstracts, Technical Reports, and Patent Disclosures or Applications:**

1. Atkinson, R. K., Miller-Talento, E., & Mallette, G. (under review). Incorporating Self-explanations and instructional explanations into Worked Examples: Does the combination capitalize on their respective strengths? American Educational Research Association, Seattle, WA.
2. Atkinson, R. K., Miller-Talento, E., & Mallette, G. (under review). Using animated agents to deliver example-based instruction: Do they support learning? American Educational Research Association, Seattle, WA.
3. Atkinson, R. K. (1999). Using a multimedia authoring systems to create innovative learning environments and support research. Paper presented at the annual meeting of the Mid-South Educational Research Association, Point Clear, AL.
4. Atkinson, R. K., & Derry, S. J. (2000). Computer-based examples designed to encourage optimal example processing: A study examining the impact of sequentially presented, subgoal-oriented worked examples. In B. Fishman & S. F. O'Connor-Divelbiss (Eds.), Proceedings of the Fourth International Conference of Learning Sciences (pp. 132-133). Hillsdale, NJ: Erlbaum.
5. Atkinson, R. K., Derry, S. D., Renkl, A., & Wortham, D. W. (in press). Learning from Examples: Instructional Principles from the Worked Examples Research. Review of Educational Research.